Bioorganic & Medicinal Chemistry Letters 26 (2016) 3594–3597

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Enriching screening libraries with bioactive fragment space

Na Zhang [a], Hongtao Zhao [b,*]

[a] College of Life Science and Bioengineering, Beijing University of Technology, Beijing 100124, China
[b] Lephar Research, Rindögatan 21, 11558 Stockholm, Sweden

ARTICLE INFO

Article history:
Received 23 May 2016
Revised 5 June 2016
Accepted 6 June 2016
Available online 7 June 2016

Keywords:
Fragment library
CK2
Kinase inhibitors
Fragment-based drug discovery

ABSTRACT

By deconvoluting 238,073 bioactive molecules in the ChEMBL library into extended Murcko ring systems, we identified a set of 2245 ring systems present in at least 10 molecules. These ring systems belong to 2221 clusters by ECFP4 fingerprints with a minimum intracluster similarity of 0.8. Their overlap with ring systems in commercial libraries was further quantified. Our findings suggest that success of a small fragment library is driven by the convergence of effective coverage of bioactive ring systems (e.g., 10% coverage by 1000 fragments vs. 40% by 2 million HTS compounds), high enrichment of bioactive ring systems, and low molecular complexity enhancing the probability of a match with the protein targets. Reconciling with the previous studies, bioactive ring systems are underrepresented in screening libraries. As such, we propose a library of virtual fragments with key functionalities via fragmentation of bioactive molecules. Its utility is exemplified by a prospective application on protein kinase CK2, resulting in the discovery of a series of novel inhibitors with the most potent compound having an IC_{50} of 0.5 μM and a ligand efficiency of 0.41 kcal/mol per heavy atom.

Drug discovery is plagued by high attrition rates at all stages,[1] urging a thorough exploration of chemical space that in principle is infinite. Complementary to the traditional high-throughput screening (HTS) where approximately 10^6 compounds will be examined against the target of interest, fragment-based drug discovery (FBDD) is gaining momentum.[2–4] It has been postulated that relatively small libraries of fragments can sample chemical space more effectively than a typical HTS library. For example, there are approximately 10^9 drug-like molecules with 11 heavy atoms or fewer,[5,6] whereas estimates vary from 10^{20} to 10^{200} for the number of guideline-of-five compliant compounds.[7,8] Nevertheless, it remains elusive why a library of 1000 fragments could have a notable hit rate against a diverse range of targets, as the chemical space for fragments typically with 8–18 heavy atoms obviously remains enormous.

In parallel to FBDD is the rise of the computational approaches. For example, the in silico fragment-based high-throughput screening cascade[9–12] has led to the discovery of inhibitors bearing novel scaffolds on the most intensively pursued kinases[9,10] and the emerging therapeutic target of bromodomains[11–13] (Table S1). Briefly, it decomposes screening libraries into fragments and then identifies in silico anchor fragments, from which parent molecules will be retrieved. Subsequently, parent molecules will be subject to molecular docking and scoring. This virtual screening approach focuses on the intrinsic fragment space, but ends with traditional low micromolar hits. Compared with the conventional docking-based high-throughput virtual screening, it can navigate the vast chemical space in an extremely efficient way.

Over a century of medicinal chemistry endeavors has led to the accumulation of hundreds of thousands of bioactive molecules across a variety of biological targets. Big data brings innovation in drug research and development in a data-driven manner. For example, probabilities for polypharmacology or off-target effects of a drug can be predicted by measuring its chemical similarity to known bioactive molecules.[14–16] By applying the in silico fragmentation approach on the 238,073 molecules accumulated in the ChEMBL library (version 12)[17] that show at least 10-μM activity on one target, we attempt to understand the phenomenal hit rate of a small fragment library by quantifying its overlap with bioactive molecules at the level of Murcko ring systems. In addition, a virtual library is proposed by fragmenting such bioactive molecules. The utility of this virtual fragment library in drug discovery was further exemplified by a prospective application on protein kinase CK2, which regulates various cellular events including signal transduction, transcriptional control, apoptosis, and cell cycle.[18] Aberrant activation of CK2 is a key oncogenic force underlying human tumorigenesis, and pharmacological inhibitors of this attractive therapeutic target have emerged as promising drug candidates.

* Corresponding author.
E-mail addresses: nanatonglei@bjut.edu.cn (N. Zhang), htzhao@lephar.com (H. Zhao).

http://dx.doi.org/10.1016/j.bmcl.2016.06.013
0960-894X/

The concept of Murcko ring systems[19] was extended to include spirocycles and atoms linked to rings via double or amide bonds. Heteroatoms are capped by methyl to keep valence intact and hence reflect their true H-bonding characteristics in molecules. The resulting ring systems represent rigid units and constitute the essential scaffold of a molecule. In addition, the size of ring systems is in the typical range of fragments, making them suitable as building blocks for rational design of fragment libraries. The 238,073 bioactive molecules in ChEMBL were deconvoluted into 8462 ring systems, among which 2573 are singletons. Similarly, the 1491 approved drugs in DrugBank[20] give rise to 485 ring systems with 231 being singletons.

A recent analysis of the attrition of drug candidates from four major pharmaceutical companies indicates a link between the physicochemical properties of compounds and clinical failure due to safety issues.[1] The molecular properties of ring systems are thus examined (Fig. S1). Briefly, the ring systems in DrugBank have an average of 12 heavy atoms, 2.3 rings, 2.3 heteroatoms (nitrogen and oxygen), and 0.43 Fsp3 (fraction of sp3 carbon). In contrast, the properties in ChEMBL change to 14 heavy atoms, 2.7 rings, 2.9 heteroatoms, and 0.41 Fsp3. The ring systems can be broken down into 163 and 1235 monocyclic, 163 and 2920 bicyclic, as well as 159 and 4307 polycyclic ring systems, in DrugBank and ChEMBL, respectively. Notably, bicyclic and in particular polycyclic ring systems become more often in ChEMBL than in DrugBank (Fig. 1). Compared with an Fsp3 of 0.62 averaged from the 4411 polycyclic ring systems in the Traditional Chinese Medicine Database (TCM),[21] the mean Fsp3 is 0.44 for the 159 polycyclic ring systems in drugs, and only 0.34 for the 4307 polycyclic ring systems in ChEMBL (Fig. S2).

Upon merging ring systems from both DrugBank and ChEMBL, we focus on a set of 2245 ring systems that occur in at least 10 molecules. These ring systems, termed as bioactive ring systems for their adequate exposure in bioactive molecules, consist of 497 monocyclic, 898 bicyclic and 850 polycyclic ring systems. They belong to 2221 clusters by ECFP4 fingerprints and a minimum intracluster similarity of 0.8. Among most common ring systems (Fig. S3), benzene appears in 84% bioactive molecules, significantly more populated than the second pyridine (10%). Highly populated ring systems often hit a variety of protein families other than a single protein. For example, heterocycles such as piperazine,[22] thiazole,[23] indole,[24] benzimidazole,[25] benzofuran,[26] benzothiazole,[27] and carbozole[28] in medicinal chemistry have been extensively reviewed.

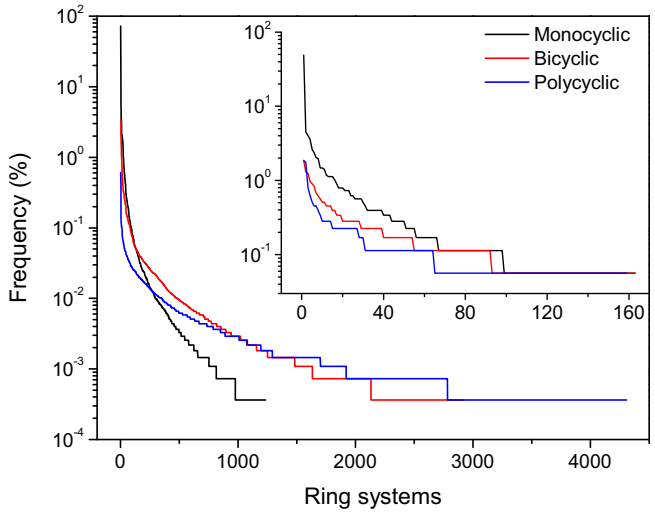

Figure 1. Pareto plot of ring systems from bioactive molecules in the ChEMBL12 library, and from drugs in DrugBank (Inset). Frequency of a ring system is the number of compounds containing the ring system divided by the size of the library.

Several commercial screening libraries were chosen to quantify their overlap with the bioactive ring systems. For HTS screening libraries from Enamine (1.9 M compounds), ChemBridge (1 M), Asinex (0.5 M), and InterBioScreen (0.5 M) to Specs (0.2 M), the number of deconvoluted ring systems shows no obvious dependency on the size of the library (Fig. 2). The InterBioScreen library is richest in ring systems (6855), followed by ChemBridge with 6792 ring systems. The 1.9-million library of Enamine corresponds to only 5780 ring systems, indicating a relatively more thorough combination of ring systems than other libraries. The proportion of the bioactive ring systems shows no obvious dependency on the library size either, with a mean percentage of 19% (Fig. 2B). With a dramatic decrease in the library size, the coverage of the bioactive ring systems gradually decreases only from 43% to 37% (Fig. 2A). Interestingly, two small HTS libraries Key organics (0.08 M) and Maybridge (0.05 M) have the highest proportion of the bioactive ring systems up to 30%, and each covers about 31% bioactive ring systems. They also have a lower proportion of polycyclic ring systems, close to that in DrugBank.

For fragment libraries with a size from 40,732 to 931 fragments, there is a significant decrease in the diversity of ring systems, and meanwhile a dramatic increase in proportion of the bioactive ring

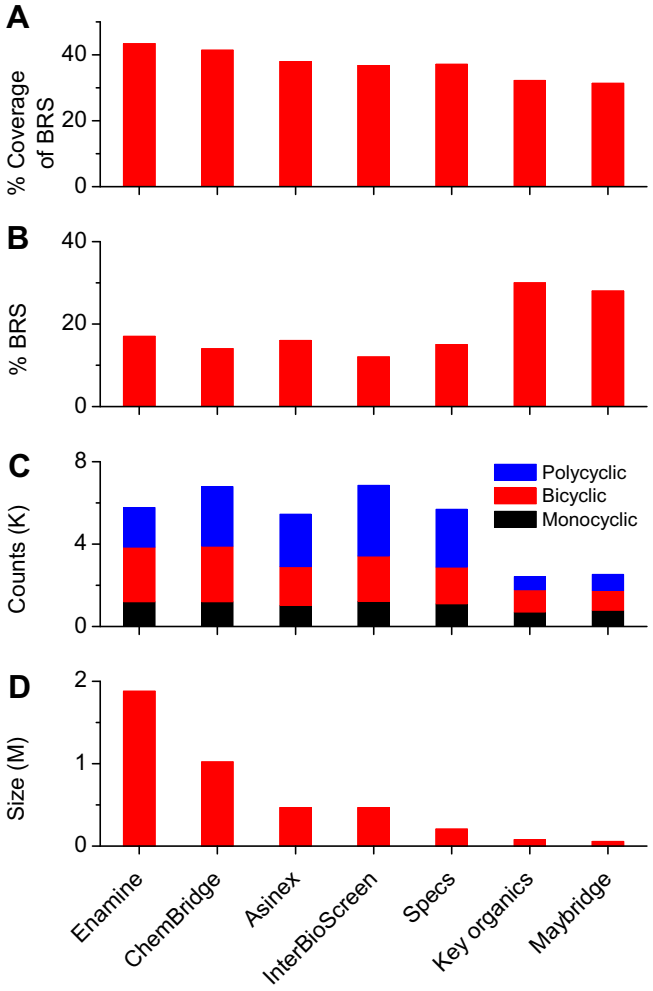

Figure 2. Overlap of the bioactive ring systems (BRS) with commercial HTS libraries. Coverage of BRS by each library (A), percentage of ring systems in each library being BRS (B), composition of ring systems in each library (C) and the library size (D).